SQL Server Consolidation on VMware Using Cisco Unified Computing System

White Paper

December 2011

Contents

Executive Summary

The purpose of this document is to provide a study of Microsoft SQL Server consolidation on Cisco Unified Compute System (UCS). The primary objective of the study is to articulate the total cost of ownership (TCO) and return on investment (ROI) that can be achieved by companies wishing to consolidate SQL Server on Cisco UCS. Additionally, the study will also prove that consolidated SQL Server implementations on Cisco UCS can meet the scalability, availability and performance requirements mandated by today's high-volume database implementations.

This white paper also provides a framework for choosing among virtualization, multi-database, and multi-instance consolidation strategies for SQL Server Database Engine supporting OLTP applications by highlighting some of the key decision points based on technical analysis.

This white paper shows the reference Architecture for considering consolidation and virtualization strategy for deploying Microsoft® SQL Server® software on the Cisco Unified Computing System (UCS).

This experiment will prove that Cisco UCS can help companies realize a reduction in total cost of ownership (TCO) and achieve the level of infrastructure agility required to meet the challenges of today's fast paced and ever changing business requirements.

Introduction

In today's economic climate, enterprises are taking a closer look within their IT organizations to identify potential areas in which cost-saving strategies can be implemented to help reduce operating expenses. One of the challenges that IT organizations face today is how to develop an infrastructure that allows flexibility, redundancy, high-availability, ease of management, security, and access control while at the same time reducing costs, hardware footprints and complexity.

The growth in IT industry has realized tremendous growth in hardware computing capacity, database applications and physical server sprawl has resulted in costly and complex computing environments containing many over-provisioned and under-utilized database servers. Many of these servers implement a single instance of SQL Server realizing ten to fifteen percent CPU utilization on average which is not an optimal use of server resources. Additionally, due to the complexities introduced by all of this growth, many database administrators today are overburdened with redundant management and administrative tasks that must be implemented on each of the many servers they are responsible for. Also, during a catastrophic, physical database server failure, there is a significant impact on the sever administrator as they typically are the resources that are required to provision another physical server into the environment and prepare it for the application.

Database server consolidation is an area where companies can realize considerable cost savings with regards to total cost of ownership (TCO). Database server consolidation can also help companies to achieve the infrastructure agility they are seeking to stay competitive and provide the fastest time to market for their solutions.

Consolidation, in general terms, is the combining of various units into more efficient and stable larger units. When applied to an IT department, consolidation specifically translates into improved cost efficiency from higher utilization of resources, standardization and improved manageability of the IT environment, and recent focus on a "green" IT environment through reduced energy consumption. One of the important components in the IT environment is the database. Databases managements systems form the foundation of many business systems. Often each group or application create their own database to solve a specific problem, thus IT departments lose control of the number of databases and servers that need to be maintained. This leads to a proliferation of databases and machines running database instances also known as database sprawl. This makes the databases the prime candidates for consolidation.

Audience and Scope

This white paper is intended for customers, partners, solution architects, storage administrators, and database administrators who are evaluating, planning a consolidation and/or virtualization strategy. It provides an overview of various considerations and reference architectures for consolidated and virtualized SQL Server 2008 R2 deployments.

Today's Challenges

IT organizations face an enormous challenge to keep operations running around the clock with increasing demand and growing complexity, to the point that it becomes difficult for employees to request resources to fulfill their tasks. This has lead to circumventing the IT standards and procedures to get the job done faster, which in turn has lead to server sprawl especially at the database tier. Database applications are typically implemented in a three four-tier environment; web-tier, application-tier and database-tier. The tiered architecture allows compartmentalization of applications and separation of resources. Hardware at each tier provides a specific function and requires a specific set of tools to manage and maintain. All of the components are connected via a network, in most cases Ethernet. This separation of resources provides an easy way for employees to bypass and deploy database servers for test and development without the IT department's oversight and management.

The apparent benefit of three-tire architectures like ease of deployment and development can easily be overshadowed by potential challenges such as incompatibility issues that may arise during deployment to production, security vulnerabilities due to different security and patching requirements, and the operating costs associated with large number of servers including support, licenses, power and cooling. Hence servers that were deployed for a specific project that is no longer economical. As the reliability and performance aspects of industry standard x86 server platforms have improved, along with the maturity of virtualization technologies on x86, many organizations are considering virtualization and consolidation for their IT.

Consideration for Consolidation

Consolidation projects are typically started to achieve specific goals such as creating room for new applications while reducing operating expenditure. These goals can be broadly grouped into the following categories:

- Standardization and centralization
- Improve floor space and power efficiency
- Reduce the number of management domains IT agility

This paper will focus on consolidation strategies for typical online transaction processing (OLTP) applications based on SQL Server.

Some general traits that make an application a good candidate for consolidation are low machine resource utilization, moderate performance requirements, little active development, and low maintenance costs. Another factor to consider is the impact on the application's network and I/O latency, because both the network and storage resources become shared as part of consolidation.

The Solution: SQL Server Consolidation Using Cisco Unified Computing System With VMware vSphere

Cisco Unified Computing System Solution Overview

Cisco Unified Computing System™ is the first converged data center platform that combines industry-standard, x86-architecture servers with networking and storage access into a single converged system. The system is entirely programmable using unified, model-based management to simplify and speed deployment of enterprise-class applications and services running in bare-metal, virtualized, and cloud-computing environments (Figure 1).

The system's x86-architecture rack-mount and blade servers are powered by Intel® Xeon® processors. These industry-standard servers deliver world-record performance to power mission-critical workloads. Cisco servers, combined with a simplified, converged architecture, drive better IT productivity and superior price/performance for lower total cost of ownership (TCO).

Building on Cisco's strength in enterprise networking, Cisco Unified Computing System is integrated with a standards-based, high-bandwidth, low-latency, virtualization-aware unified fabric . The system is wired once to support the desired bandwidth and carries all Internet protocol, storage, inter-process communication, and virtual machine traffic with security isolation, visibility, and control equivalent to physical networks. The system meets the bandwidth demands of today's multicore processors, eliminates costly redundancy, and increases workload agility, reliability, and performance.

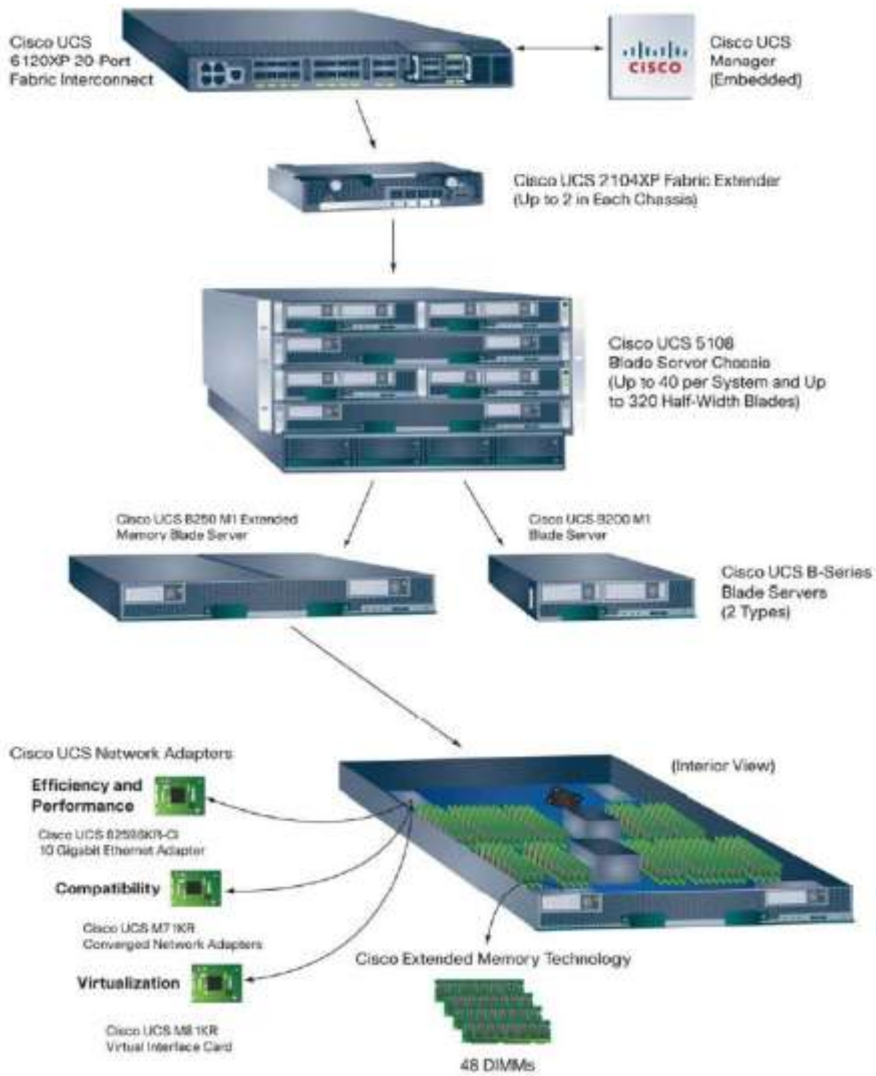
Cisco Unified Computing System is designed from the ground up to be programmable and self integrating. A server's entire hardware stack, ranging from server firmware and settings to network profiles, is configured through model-based management. With Cisco virtual interface cards, even the number and type of I/O interfaces is programmed dynamically, making every server ready to power any workload at any time.

With model-based management, administrators manipulate a model of a desired system configuration, associate a model's service profile with hardware resources, and the system configures itself to match the model. This automation speeds provisioning and workload migration with accurate and rapid scalability. The result is increased IT staff productivity, improved compliance, and reduced risk of failures due to inconsistent configurations.

Cisco Fabric Extender technology reduces the number of system components to purchase, configure, manage, and maintain by condensing three network layers into one. It eliminates both blade server and hypervisor-based switches by connecting fabric interconnect ports directly to individual blade servers and virtual machines. Virtual networks are now managed exactly as physical networks are, but with massive scalability. This represents a radical simplification over traditional systems, reducing capital and operating costs while increasing business agility, simplifying and speeding deployment, and improving performance.

Cisco Unified Computing System helps organizations go beyond efficiency: it helps them become more effective through technologies that breed simplicity rather than complexity. The result is flexible, agile, high-performance, self-integrating information technology, reduced staff costs with increased uptime through automation, and more rapid return on investment.

Figure 1. Cisco Unified Computing System Components



VMware vSphere 4.1 Architecture

VMware vSphere 4.1 Hypervisor is based on the VMware ESXi hypervisor architecture and as such it inherits all the new capabilities that VMware introduced with the 4.1 release of ESX.

VMware ESX are “bare-metal” hypervisor architectures, meaning they install directly on top of the physical server and partition it into multiple virtual machines that can run simultaneously, sharing the physical resources of the underlying server. Each virtual machine represents a complete system, with processors, memory, networking, storage and BIOS, and can run an unmodified operating system and applications.

Microsoft SQL Server 2008 R2

SQL Server 2008 R2 Enterprise delivers a comprehensive data platform that provides built-in security, availability, and scale coupled with robust business intelligence offerings—helping enable the highest service levels for mission-critical workloads. It is designed to deliver a high-performing data platform that provides the highest levels of scalability for large application workloads, virtualization and consolidation, and management for an organization’s database infrastructure. Datacenter helps enable organizations to cost effectively scale their mission-critical environment.

Microsoft SQL Server 2008 R2 is a complete set of enterprise-ready technologies and tools that help organizations derive the most value from information while lowering total cost of ownership (TCO). With Microsoft SQL Server 2008 R2, organizations can:

- Implement peer-to-peer replication quickly with the new visual designer and add nodes without affecting system availability
- Use Microsoft Performance Studio to troubleshoot, tune, and monitor Microsoft SQL Server 2008 instances across the enterprise
- Take control of workload resource use with Microsoft Resource Governor
- Reduce storage requirements and improve performance with data compression, backup compression, sparse columns, and filtered index and backup compression
- Optimize database mirroring performance and eliminate downtime with automatic recovery of suspect pages

Key features:

- Application and Multi-Server Management for enrolling, gaining insights and managing over 25 instances
- Highest virtualization support for maximum ROI on consolidation and virtualization
- Supports more than 8 physical processors for highest levels of scale
- Supports memory limits up to OS maximum

EMC VNX Storage System

The EMC® VNX™ family of storage systems represents EMC’s next generation of unified storage optimized for virtualized environments. The massive virtualization and consolidation trend of servers demands a new storage technology that is dynamic and scalable. The EMC VNX series offers several software and hardware features for optimally deploying mission-critical enterprise applications.

The new generation of EMC unified storage offers a range of choices for meeting the diversified business requirements of the enterprise, including performance, capacity, and protection, at the best total cost of ownership.

Figure 2. The EMC VNX Family of Unified Storage Platforms

A key distinction of this new generation of platforms is support for both block- and file-based external storage access over a variety of access protocols, including Fibre Channel (FC), iSCSI, FCoE, NFS, and CIFS network shared file access. Furthermore, data stored in one of these systems, whether accessed as block- or file-based storage objects, is managed uniformly via Unisphere, a web-based interface window. Additional information on Unisphere can be found on emc.com in the white paper titled: Introducing EMC Unisphere: A Common Midrange Element Manager.

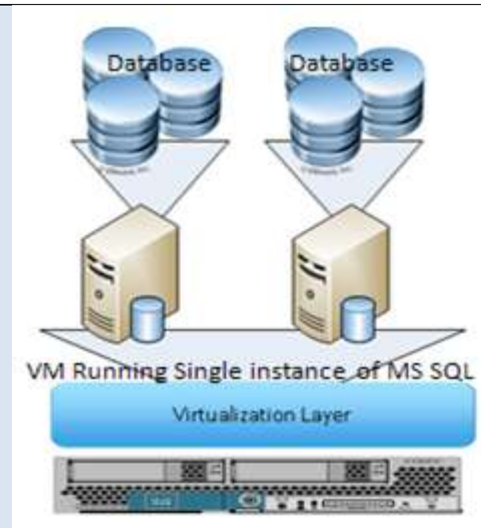
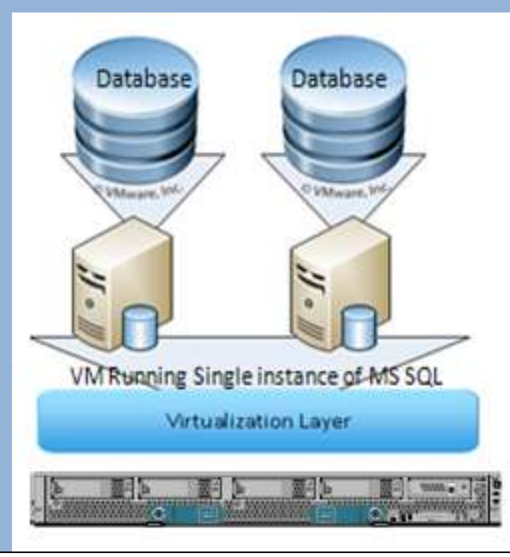
EMC's new VNX storage family now supports the 2.5" SAS drives in a 2U disk array enclosure (DAE) that can hold up to 25 drives, one of the densest offerings in the industry. For example, compared to the older-generation technology of storing 15 x 600 GB worth of data using the 3.5" FC drives in a 3U DAE, the new DAE using 25 x 600 GB drives in a 2U footprint means an increase by 2.5 times. The power efficiency of the new DAEs also makes it more cost-effective to store the increased data in this much more compact footprint without the need to increase power consumption and cooling.

Server Consolidation Case Study

Case Study: Microsoft SQL Server 2008 R2 Using VMware vSphere 4.1

This section details the scenario implemented for this solution. The SQL Server testing is carried out in four different Cisco UCS platform (B200 M2, B230 M2, B250 M2 and B440 M2) using VMware vSphere 5 by considering the following two potential scenarios (Figure 3).

Figure 3. Microsoft SQL Server Deployment on VMware and Cisco UCS Server Blades

<table>
<tr>
<td>Multiple virtual machines per physical server hosting multiple small databases in each VM</td>
<td>
</td>
</tr>
<tr>
<td>Multiple virtual machines per physical server hosting single large database in each VM</td>
<td>
</td>
</tr>
</table>

Configuration 1: Using a single physical machine to host multiple virtual machines and each virtual machine running a Microsoft SQL Server instances hosting multiple Small & Medium databases.

Configuration 2: Using a single physical machine to host multiple virtual machines and each virtual machine running Microsoft® SQL Server® database instance hosting a single Large database.

Each of these scenarios have their own security, compliance requirements, high availability, disaster recovery requirements, resource management benefits, level of consolidation density, and manageability characteristics.

Both these options have been tested in this study. For Configuration 1 we used a small database of 10 GB size and for Configuration 2 we used a large database of 100 GB size.

By considering these two consolidation configurations, we divided the testing in three different scenarios:

Scenario 1: SQL Server consolidation using Cisco UCS B200 M2, B230 M2 with 10 GB database.

In this Scenario, we choose Cisco UCS B200 M2 & Cisco UCS B230 M2 blades for small database consolidation. In fact, Cisco UCS B200 M2 is a entry level blade server from Westmere-EP platform family, supports max. up to 192GB of Physical Memory, which can host small databases with relatively low disk bandwidth and Cisco UCS B230 M2 is a entry level server from Westmere EX Platform family, which supports 256 GB of Physical Memory, which is also ideal candidate for small database consolidation.

Scenario 2: SQL Server consolidation and Scalability testing using Cisco UCS B250 M2 with 10 GB and 100 GB Databases.

In this Scenario, we choose Cisco UCS B250 M2 blade server for small database and large database consolidation. The Cisco UCS B250 M2 Extended Memory blade server uses patented Cisco Extended Memory Technology. This Cisco technology provides more than twice as much industry-standard memory (384 GB) as traditional two-socket servers, increasing performance and capacity for demanding virtualization and large-data-set workloads. This B250 M2 Extended Memory blade Server can be a high end for small database consolidation and entry level for large Database consolidation. in a Virtualized Environment.

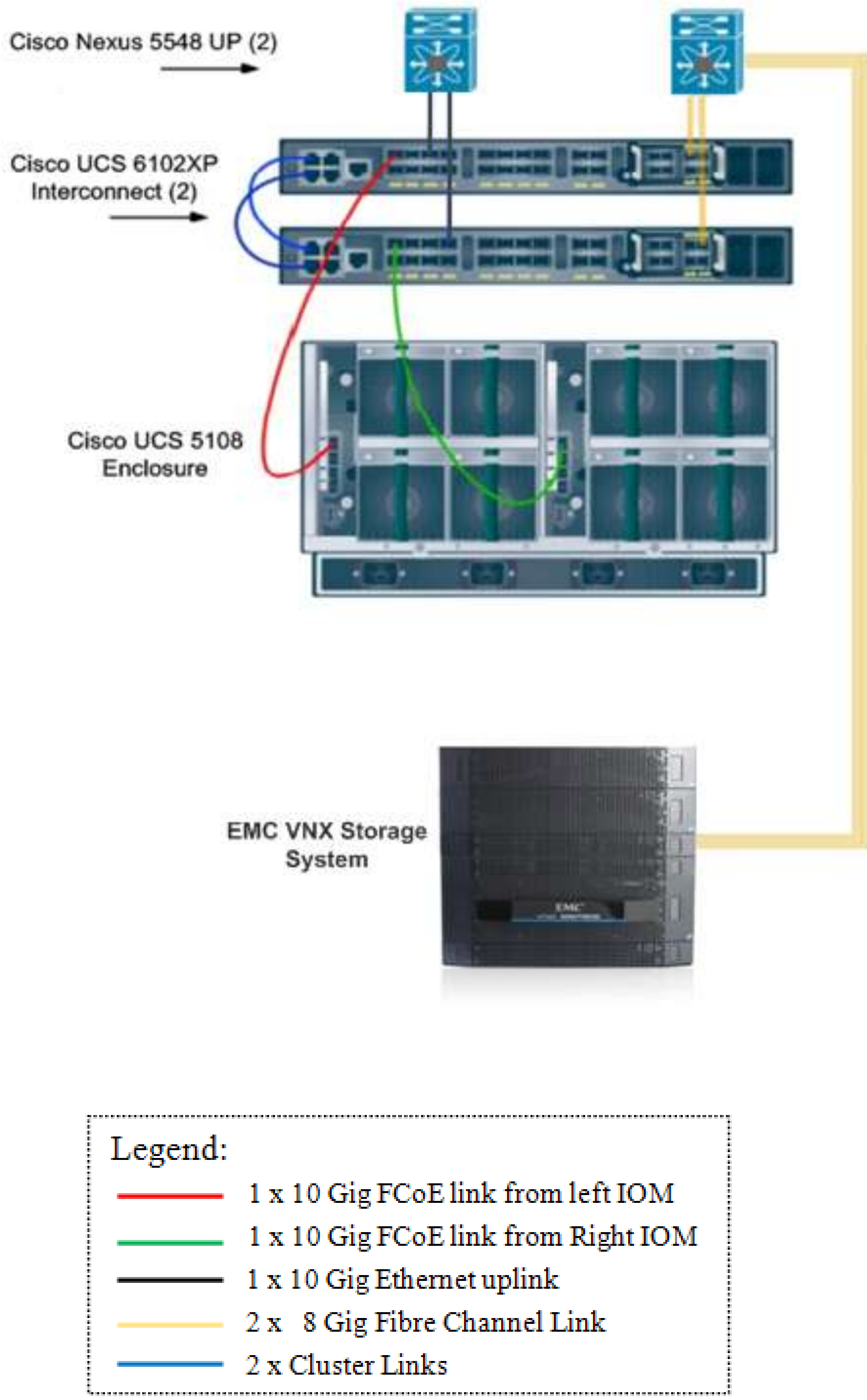
Scenario 3: SQL Server consolidation using Cisco UCS B440 M2 with 100 GB database

In this Scenario, we choose Cisco UCS B440 M2 blade server for large database consolidation. This Cisco UCS B440 M2 is a 4-socket blade server from Westmere EX platform family, supports up to 40 Physical cores and 512 GB of Physical Memory, which is a high end Blade server for large Database Consolidation.

Testing Methodology and Configurations

Figure 4 details the System Under Test (SUT).

Figure 4. System Under Test



Test Configurations

Table 1 shows the Cisco UCS Platform used and tested in the Study.

Table 1. Cisco Unified Computing System Reference Configuration

HARDWARE SPECIFICATION	CISCO UCS B200 M2 BLADE SERVER	CISCO UCS B230 M2 BLADE SERVER	CISCO UCS B250 M2 BLADE SERVER	CISCO UCS B440 M2 BLADE SERVER
CPU Model	Intel Xeon X5690	Intel Xeon E7-2870	Intel Xeon X5690	Intel Xeon E7-4870
CPU Speed	3.46 GHz	2.40 GHz	3.46 GHz	2.40 GHz
Number of Processors	2	2	2	4
Number of Cores per Processor	6	10	6	10
Number of Hardware threads per core	2	2	2	2
Memory Type	DDR3 PC3-10600R	DDR3 PC3-10600R	DDR3 PC3-10600R	DDR3 PC3-10600R
Total Memory	192 GB (12 x 16 GB)	256 GB (32 x 8 GB)	384 GB (12 x 8 GB)	512 GB (32 x 16 GB)

Table 2 shows the settings used on Virtual Machines on all four test servers. In this test, in general, the more RAM the blade server can support more virtual machines the blade server can run.

Table 2. Virtual Machine Configuration for Small and Large Database Consolidation Study

VMWARE MACHINE SPECIFICATIONS	VIRTUAL MACHINE SETTINGS (FOR SMALL DATABASE)	VIRTUAL MACHINE SETTINGS (FOR LARGE DATABASE)
Virtual CPU (vCPU)	4	8
Memory	32 GB	64 GB
Virtual NIC (vNIC) Type	VMXNET 3	VMXNET 3
Number of Virtual Disks	3 (OS, SQL Logs and SQL Database)	3 (OS, SQL Logs and SQL Database)

Table 3 shows the EMC Storage Configurations during SQL Study on each Cisco UCS Platform.

Table 3. EMC Storage Configurations – SQL Study

STORAGE SPECIFICATION	CISCO UCS B200 M2 BLADE SERVER	CISCO UCS B230 M2 BLADE SERVER	CISCO UCS B250 M2 BLADE SERVER	CISCO UCS B440 M2 BLADE SERVER
Number of DAE Enclosures Used	3	3	4 - Small DB & 3 - Large DB	3
Number of Disks Used for OS	6 (RAID 10 (3+3))	6 (RAID 10 (3+3))	6 (RAID 10 (3+3))	6 (RAID 10 (3+3))
Number of Disks Used for SQL LOG	8 (RAID 10(4+4))	8 (RAID 10(4+4))	8 (RAID 10(4+4))	8 (RAID 10(4+4))
Number of Disks Used for SQL Data	15 (RAID 5(4+1) x 3)	20 (RAID 5(4+1) x 4)	30 (RAID 5(4+1) x 6) – Small DB & 15 (RAID 5(4+1) x 3) – Large DB	15 (RAID 10(4+1) x 3)

Logical Layout of Tested Configurations

Scenario 1: SQL Server consolidation using Cisco UCS B200 M2, B230 M2 with 10 GB database

Figure 5 shows the ogical layout of SQL server consolidation in a virtualized environment using Cisco UCS B200 M2 blade server with small database (10 GB).

Figure 5. Logical Layout of Small Database Consolidation Using Cisco UCS B200 M2

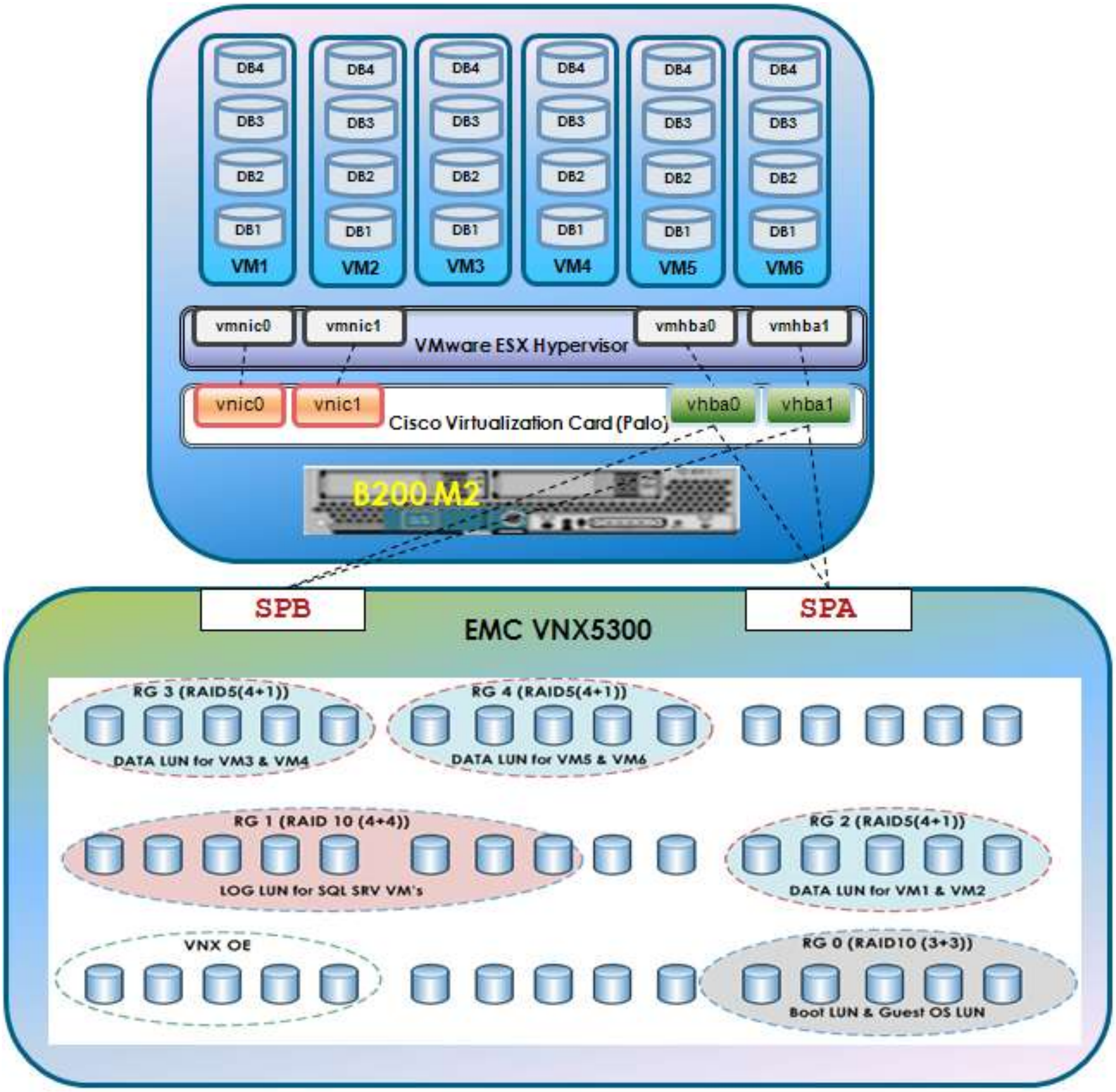


Figure 6 shows the logical layout of SQL server consolidation in a virtualized environment using Cisco UCS B230 M2 blade server with small database (10 GB).

Figure 6. Logical Layout of Small Database Consolidation Using Cisco UCS B230 M2

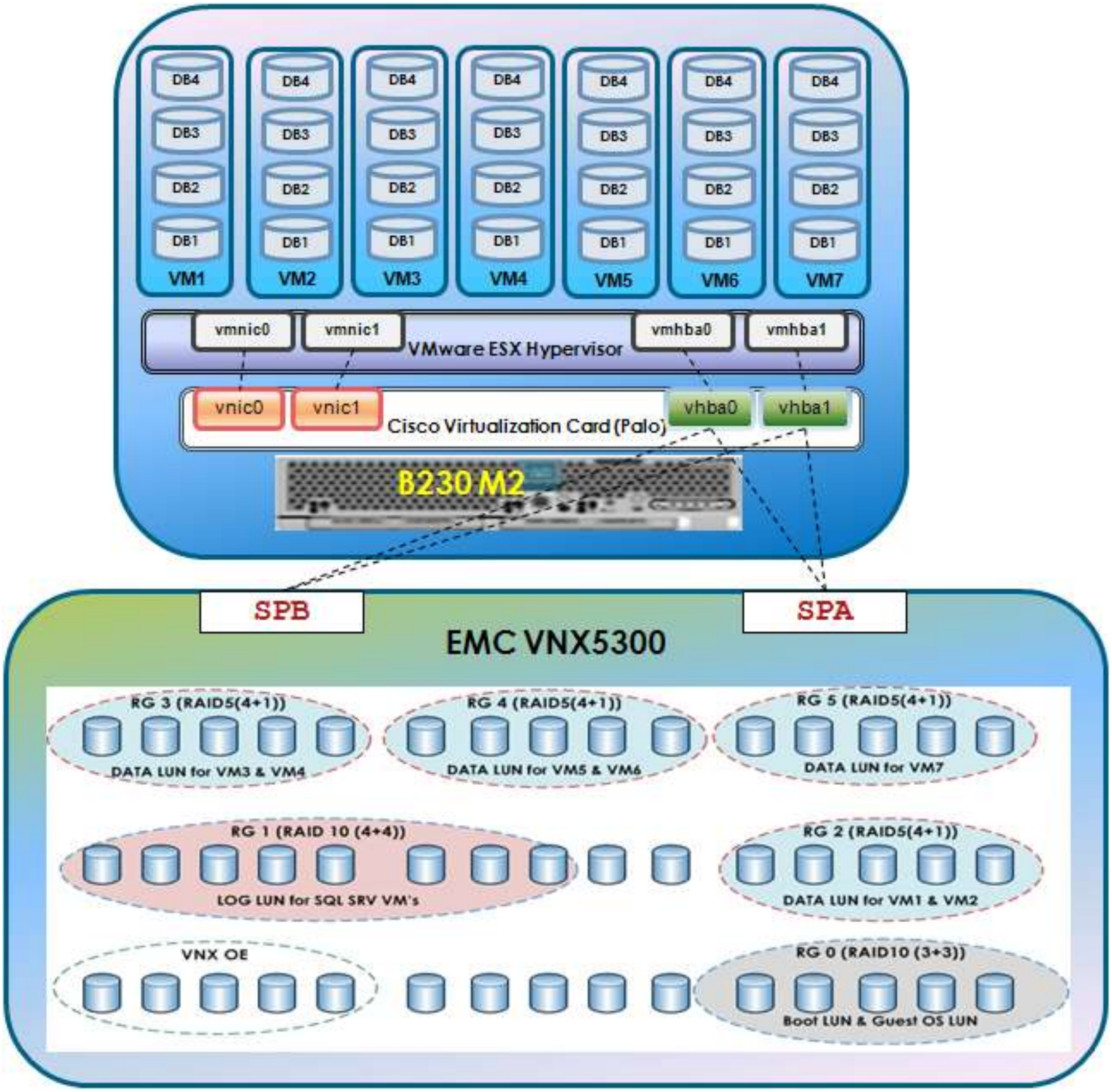


Figure 7 shows the logical layout of SQL server consolidation in a virtualized environment using Cisco UCS B250 M2 blade server with small database (10 GB).

Figure 7. Logical Layout of Small Database Consolidation Using Cisco UCS B250 M2

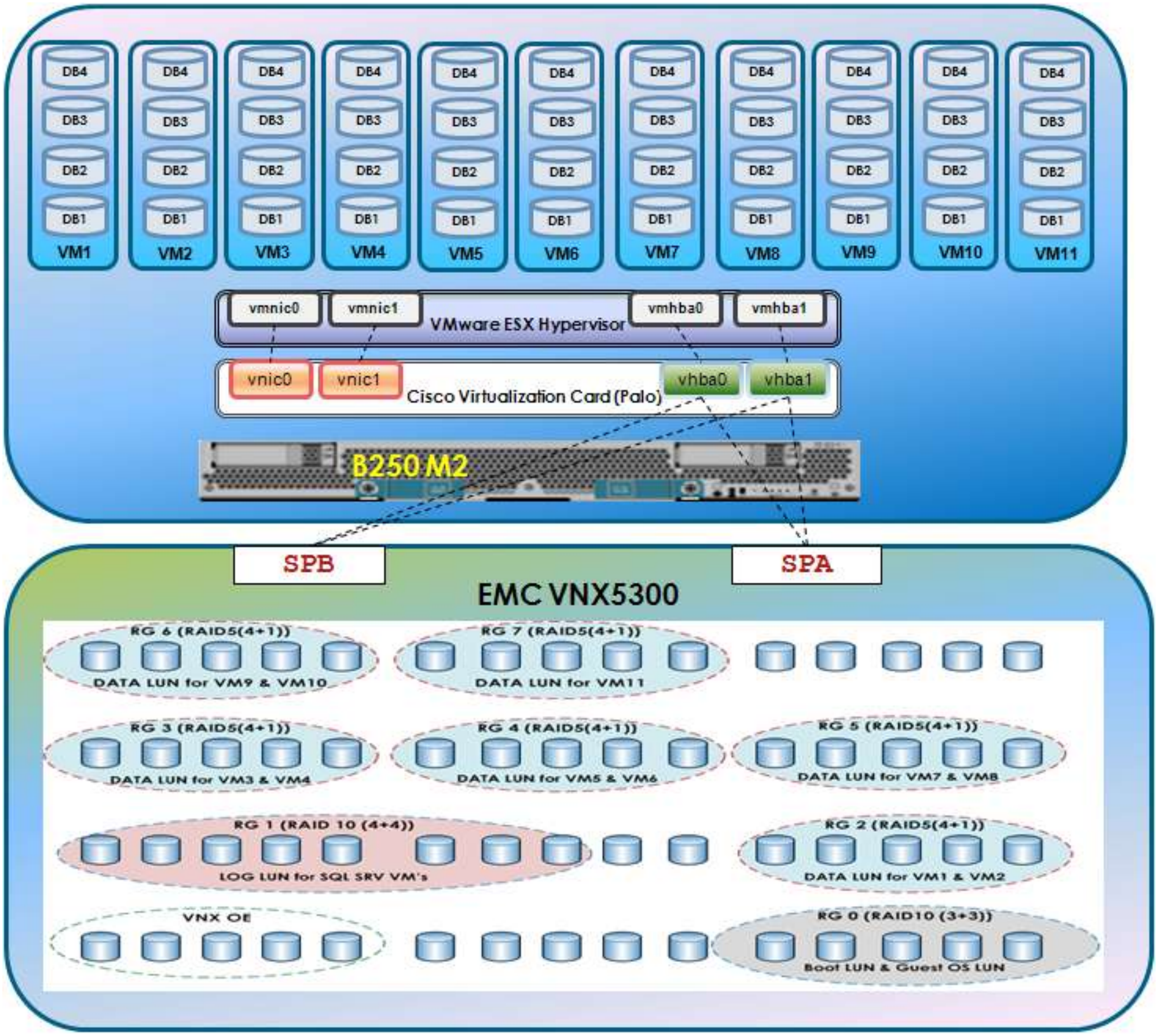


Figure 8 shows the logical layout of SQL server consolidation in a virtualized environment using Cisco UCS B250 M2 with large database (100 GB).

Figure 8. Logical Layout of Large Database Consolidation Using Cisco UCS B250 M2

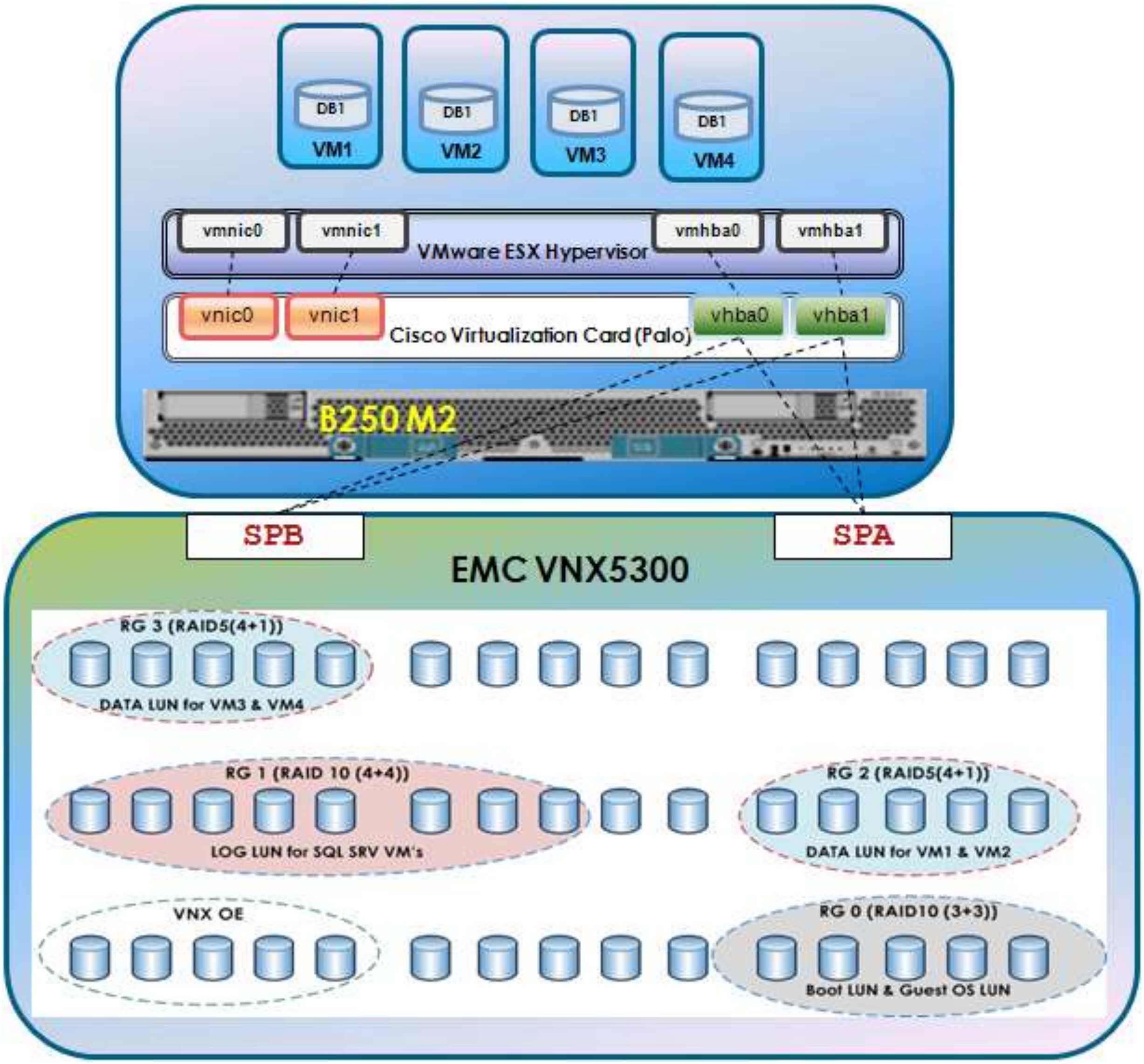
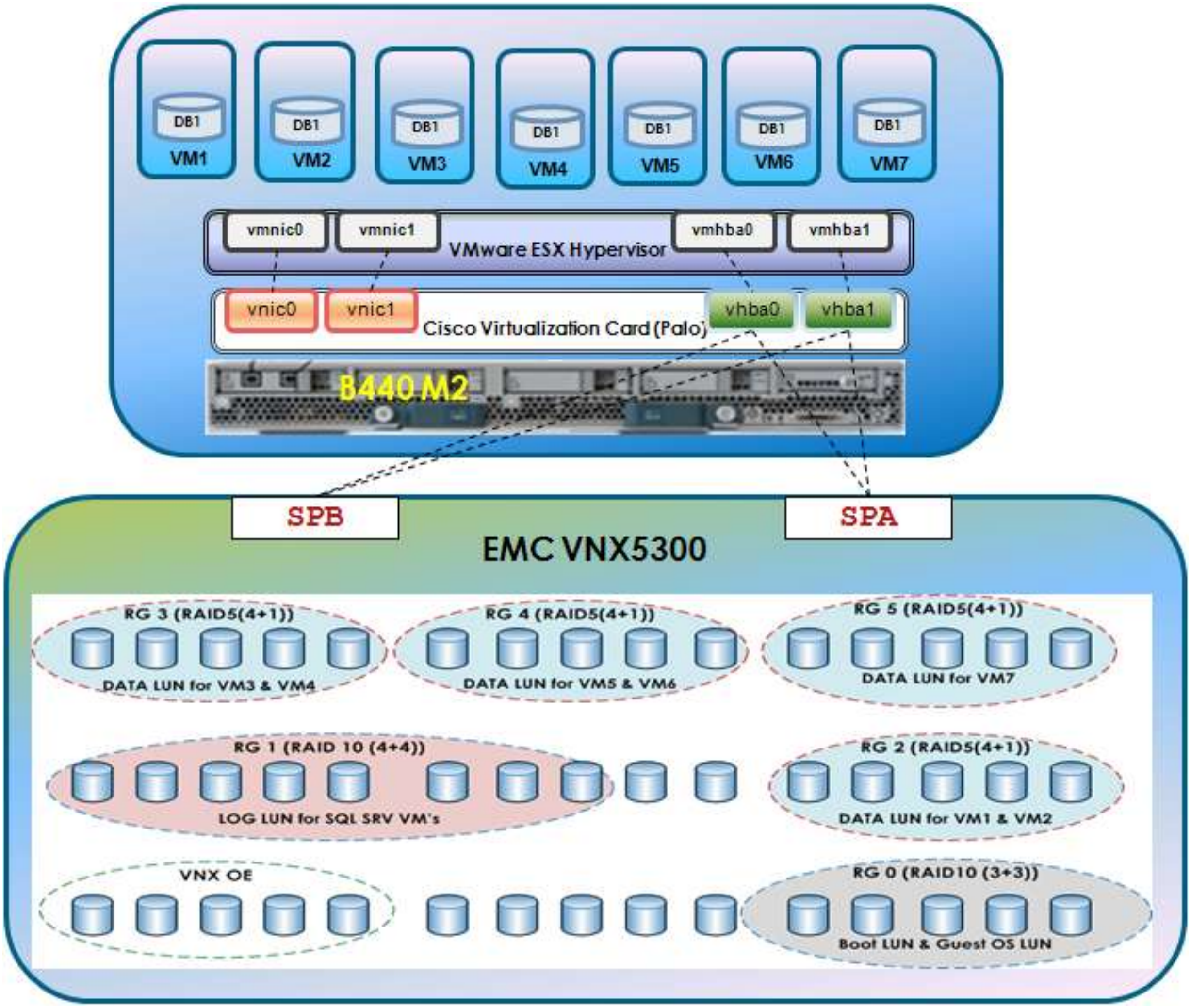


Figure 9 shows the logical layout of SQL server consolidation in a virtualized environment using Cisco UCS B440 M2 using small database (100 GB).

Figure 9. Logical Layout of Large Database Consolidation Using Cisco UCS B440 M2



Workload

To build the workload, we used a typical OLTP transaction involving ordering, searching and delivery. We have used different parameters such as think time, thread etc. for two different sizes of database to saturate each VIRTUAL MACHINE. We tested with two different sizes of databases in two different scenarios, in this consolidation process. In the first scenario we created small database (for small and medium scale business) and up to four databases in one SQL Server instance running on each virtual machine. In second scenario we created large database (for large scale business) and one copy of large size database in one SQL server Instance running on each virtual machine. Client machine was running windows 2003 to initiate the workload for each database.

In the first scenario, we started the test with one virtual machine and one small database in one SQL server instance captured the benchmark score and resource utilization. Later we started adding one small database at a time to the same SQL server instance on the same virtual machine and simultaneously started the workload on all the databases on that virtual machine. In this way we tested up to four databases on the same virtual machine and captured the benchmark score at every instance of adding new database to the virtual machine. It was found throughput scaled linearly up to four databases in one single SQL Server instance on one virtual machine.

Following that we then added more virtual machines to the configuration with each virtual machine now supporting four databases under one SQL server instance. Benchmark scores and resource usages were collected with this configuration as well.

In the second scenario, we started the test with one database (large database) in one SQL instance on one virtual machine and captured the benchmark score and resource utilization. Later we started adding one virtual machine at a time to the same physical machine and simultaneously started the workload on all the virtual machines. In this way, more virtual machines added to the physical machine and the workload has been tested and captured the benchmark score at every instance of adding new virtual machine.

The primary metric is OPM, which the driver program calculates and reports in a test file on the client machines.

Table 5 details the workload parameters used for 10 GB and 100 GB consolidation study.

Table 4. OLTP Client Parameters for Small and Large Database Consolidation Study

OLTP PARAMETER DESCRIPTION	PARAMETER VALUE FOR SMALL DATABASE	PARAMETER VALUE FOR LARGE DATABASE
Database Size (GB)	10 GB	100 GB
Number of Threads	32	32
Think time (sec)	0.3 sec	0.001 sec
Startup rate (users/sec)	10	10
Run time(min)- 0 is Infinite	15 min	15 min
Warmup time(min)	15 min	15 min

Test Results

Scenario 1: SQL Server Consolidation Using Cisco UCS B200 M2 and B230 M2 With 10 GB DB

Figure 10 shows the resource utilization of Cisco UCS B200 M2, especially CPU and memory of the physical Machine after adding each virtual machine one after another with four small databases in each virtual machine and concurrently running OLTP workload on each virtual machine. Workload was capped to make sure that resources (Memory and CPU) were not saturated. In this testing process we observed 93 percent of memory utilization and 67 percent CPU utilization. As it is shown in the graph, physical machine still had enough CPU to handle any unexpected spikes in the workloads.

Figure 10. CPU and Memory Utilization of Physical Machines After Adding Each Virtual Machine

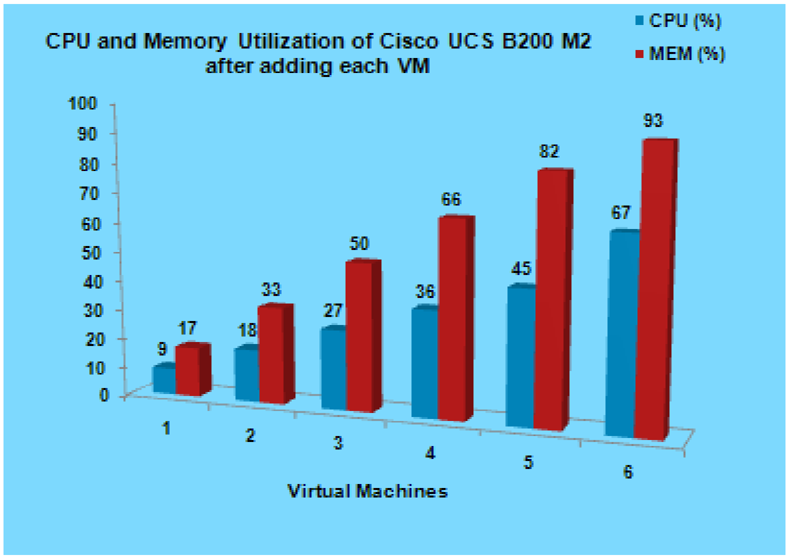


Figure 11 shows the results of Operations per Minute (OPM), the OPM has been recorded after adding 4 databases on each virtual machine and there were six virtual machines created on the same physical machine. We found linear scaling of OPM as well as the linear scaling of resource utilization. In the end, while testing 24 databases on six virtual machines, OPM tapered due to 93 percent of memory consumption.

Figure 11. OPM and Database Scaling

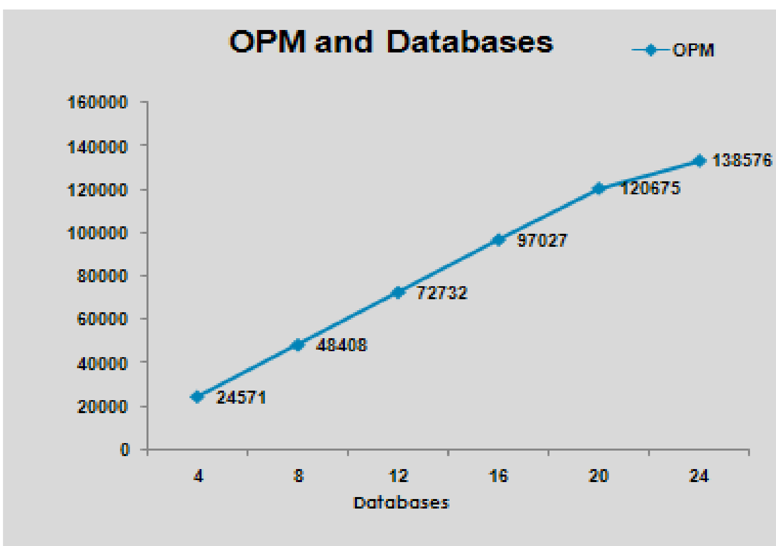


Figure 12 shows the IO characteristics of Cisco UCS B200 M2, the IOPs has been measured using EMC Unisphere monitoring GUI interface and it is recorded after adding each virtual machine into the same physical machine. In this scenario we scaled up to 6 virtual machines on Cisco UCS B200 M2, which typically uses 15 disks for data lun configured with RAID 5 ((4+1)*3) and 8 disks LUN for Log configured with RAID 10(4+4). Note that there is enough capacity both from fabric as well the storage perspective to support more IOPS and large bandwidth. This is because workload is memory bound than I/O bound.

Figure 12. IO Characteristics of Cisco UCS B200 M2 Measured After Adding Each Virtual Machine

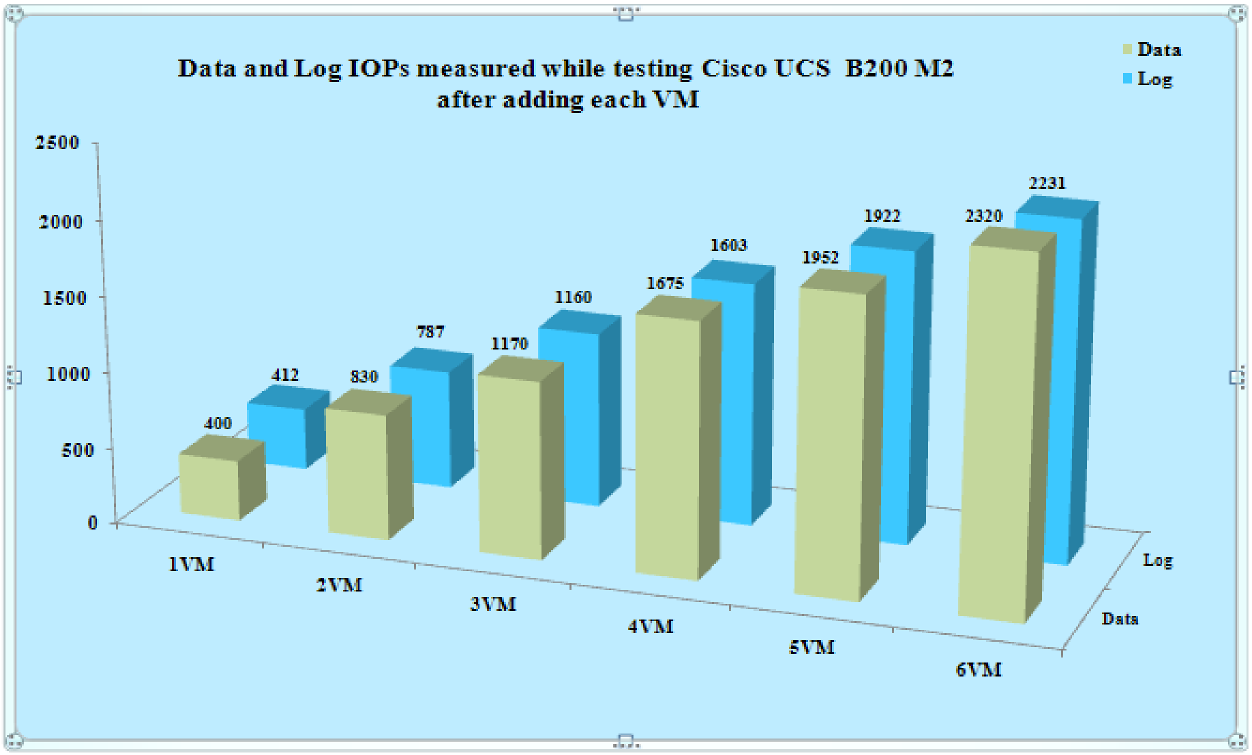


Figure 13 shows the test results of Cisco UCS B200 M2, Transaction per Minute (TPM) is measured using Performance monitor and it is recorded after adding 4 databases on each virtual machine into the same physical machine. Since, the physical server ran out of memory while six virtual machines were running, the TPM started tapering down.

Figure 13. Transaction Per Minute (TPM) Scaling After Adding Each Virtual Machine

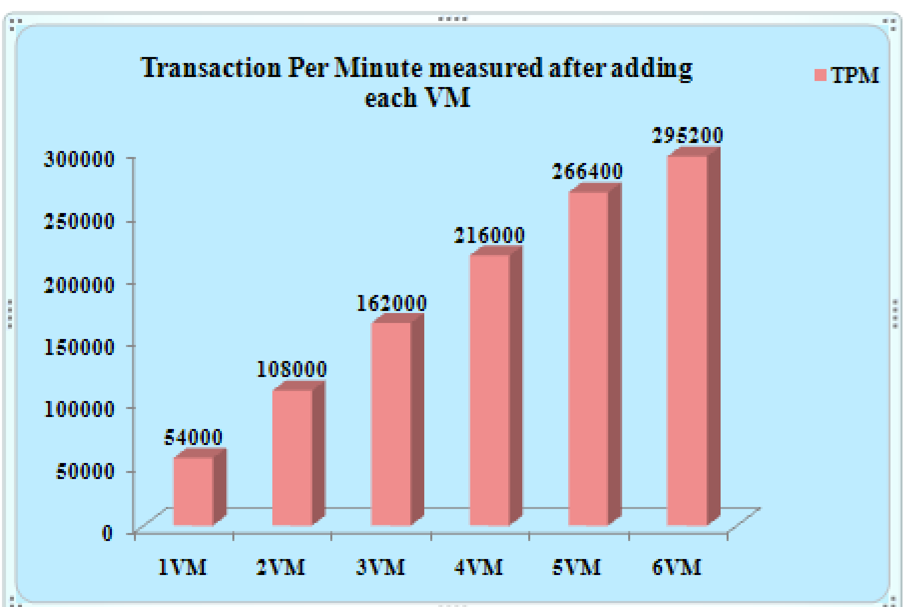


Figure 14 shows the resource utilization of Cisco UCS B230 M2, especially CPU and memory of physical machine after adding each virtual machine one after another with four small databases running on each virtual machine and concurrently running OLTP workload on each virtual machine. The workload was capped to make sure that resources (Mem and CPU) were not saturated. In this testing process we observed 87 percent of memory utilization and 50 percent of CPU utilization that means the physical server has plenty of CPU resources left and by adding more memory to the server, we can scale more than seven virtual machines. Note that with higher memory (up to 512G) we can easily support upto 17 virtual machines (68 databases). This is evident with the CPU being utilized 50 percent while memory is 87 percent for 256G configuration.

Figure 14. CPU and Memory Utilization of Physical Machines After Adding Each Virtual Machine

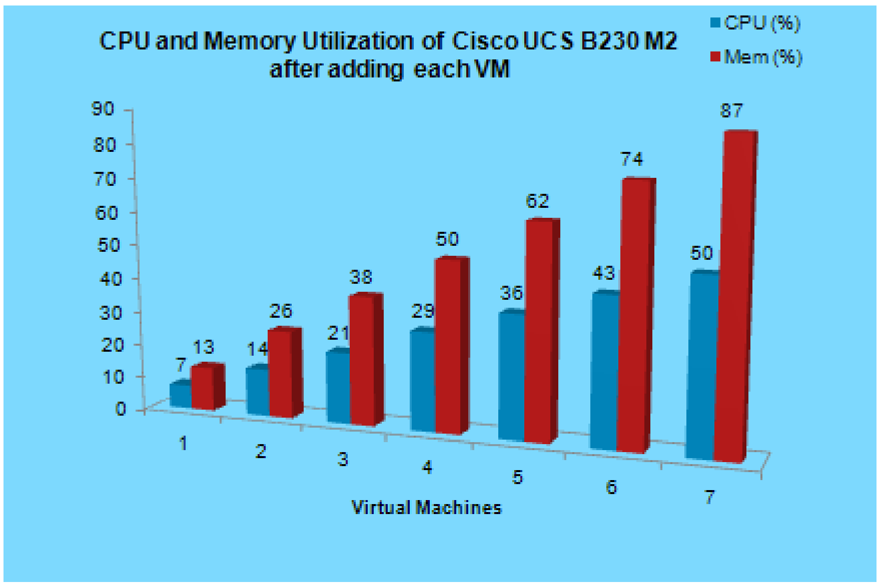


Figure 15 shows the test results of Cisco UCS B230 M2, OPM recorded after adding four databases each virtual machine into the same physical machine. In this scenario we added 4 databases to each virtual machine and there were seven virtual machine created on the same physical machine. We found linear scaling of OPM as well as the linear scaling of resource utilization.

Figure 15. OPM and Database Scaling

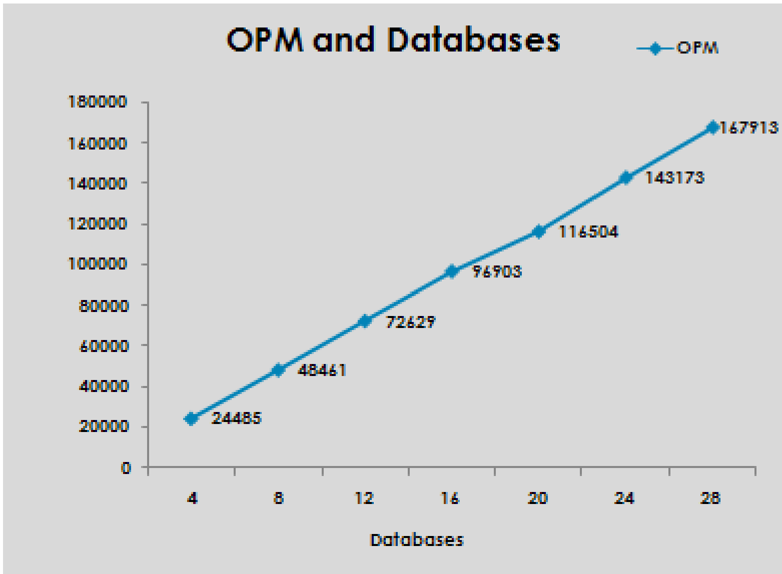


Figure 16 shows the IO characteristics of Cisco UCS B230 M2, the IOPs has been measured using EMC Unisphere monitoring GUI interface and it is recorded after adding each virtual machine into the same physical machine. In this scenario we scaled up to 7 virtual machines on Cisco UCS B230 M2, which typically uses 20 disks for data lun configured with RAID 5 ((4+1)*4) and 8 disks LUN for Log configured with RAID 10(4+4).

Figure 16. IO Characteristics of Cisco UCS B230 M2 Measured After Adding Each Virtual Machines

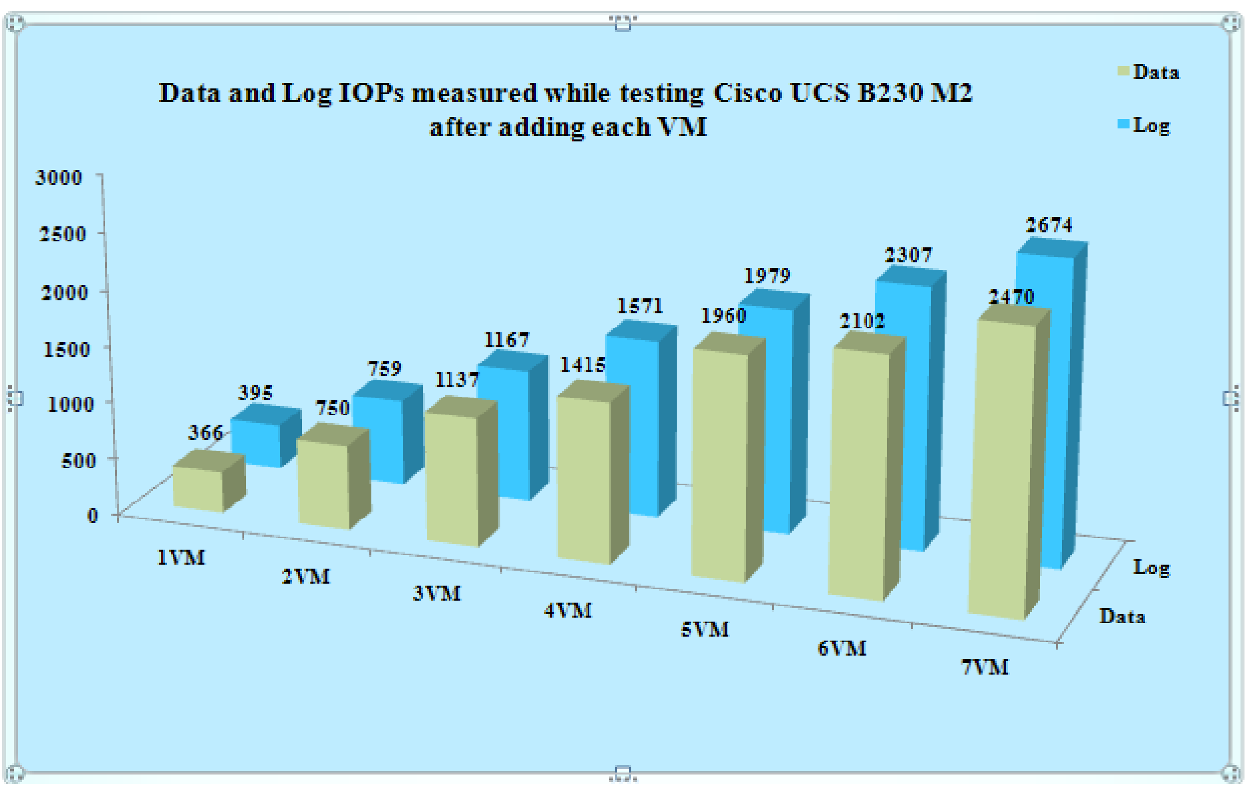
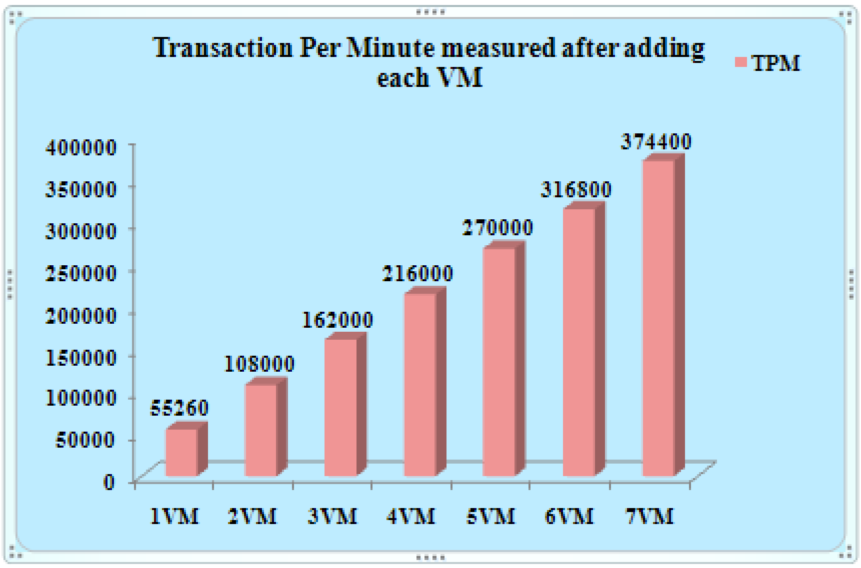


Figure 17 shows the test results of Cisco UCS B230 M2, Transaction per Minute (TPM) is measured using Performance monitor and it is recorded after adding 4 databases on each virtual machine into the same physical machine. In this scenario we added 4 databases to each virtual machine and there were 7 virtual machines created on the same physical machine.

Figure 17. Transaction per Minute Scaling After Adding Each Virtual Machine



Scenario 2: SQL Server Consolidation and Scalability Testing Using Cisco UCS B250 M2 with 10 GB and 100 GB Databases

Figure 18 shows the resource Utilization of Cisco UCS B250 M2, especially CPU and Memory resources measured using esxtop tool after adding each Virtual Machine one after another with 4 small databases running on each virtual machine and concurrently running OLTP workload on each virtual machine. In this testing process we observed 88 percent of memory utilization and 98 percent CPU Utilization on the Physical server. As shown in the graph, after scaling upto 11 virtual machines, the physical server CPU resources saturate and still left with more memory resources. The reason for CPU saturation is due to more number of virtual CPU's (vCPU) created on the Cisco UCS B250 M2 server, which supports upto 12 physical processor cores and 12 logical cores with Hyper threading enabled. Each virtual machine created with 4 vCPU and 32GB of memory, in total we created upto 11 virtual machines, using 44 virtual CPU's and 352 GB of memory from the Physical machine. Workload rendered the system CPU bound and near saturation of memory.

Figure 18. CPU and Memory Utilization of Physical Machines After Adding Each Virtual Machine

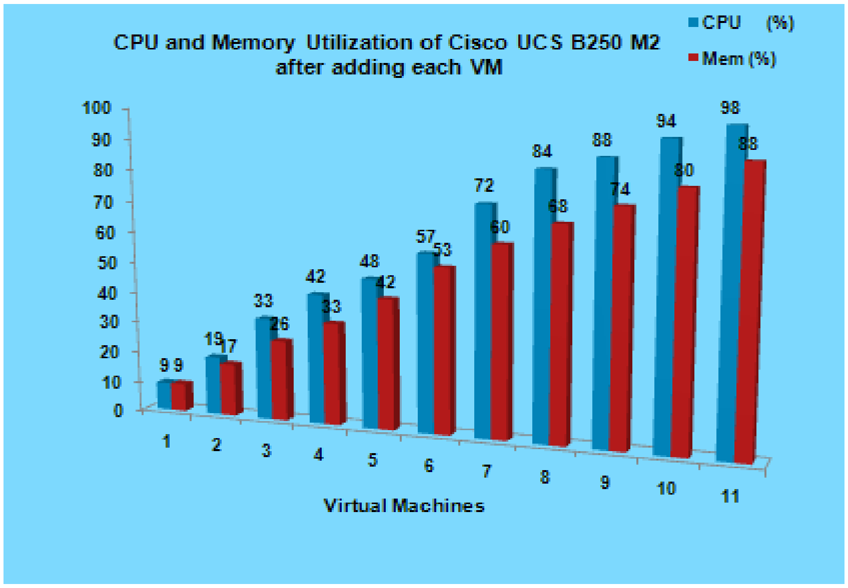


Figure 19 shows the test results of Cisco UCS B250 M2, OPM recorded after adding 4 databases on each virtual machine into the same physical machine. In this scenario we added 4 databases to each virtual machine and there were 11 virtual machines created on the same physical machine. We found linear scaling of OPM as well as the linear scaling of resource utilization.

Figure 19. OPM and Database Scaling

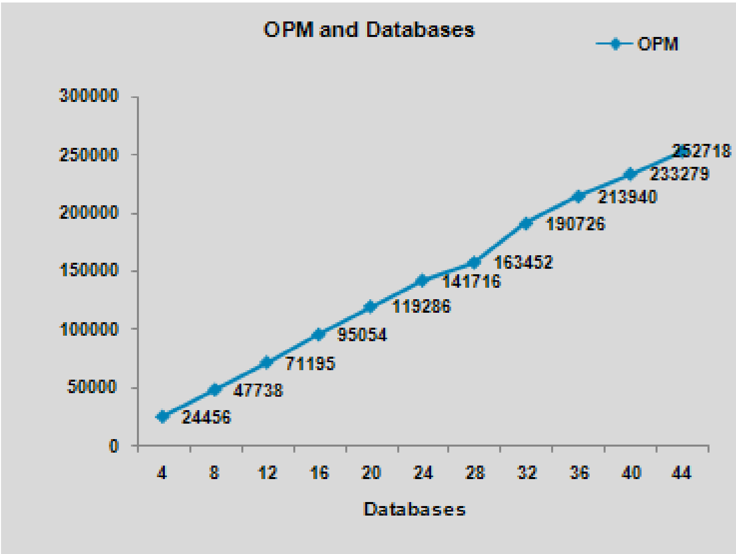


Figure 20 shows the IO characteristics of Cisco UCS B250 M2, the IOPs has been measured using EMC Unisphere monitoring GUI interface and it is recorded after adding each virtual machine into the same physical machine. In this scenario we scaled up to 11 virtual machines on Cisco UCS B250 M2, which typically uses 30 disks for data lun configured with RAID 5 ((4+1)*6) and 8 disks for Log LUN configured with RAID 10(4+4). As shown in the graph, the I/O throughput drops down, due to physical server CPU saturation, which causes the system to generate more I/O's and it is not because of the I/O bandwidth or limitation.

Figure 20. IO Characteristics of Cisco UCS B250 M2 Measured After Adding Each Virtual Machine

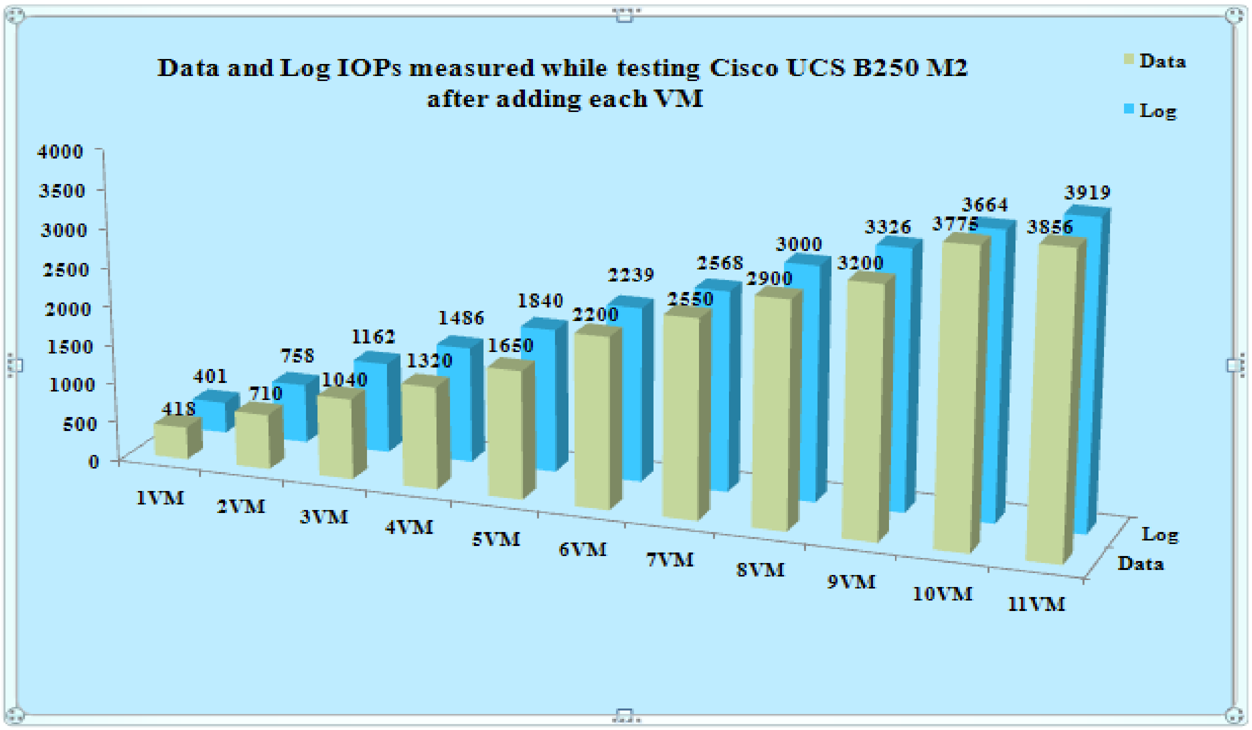


Figure 21 shows the results of TPM, the Transaction per Minute (TPM) is measured using Performance monitor on each virtual machine and it is recorded after adding 4 databases on each virtual machine and there were 11 virtual machines created on the same physical machine. Total of 44 databases were hosted on single B250 M2 machine with sustained throughput and response time.

Figure 21. Transaction per Minute Scaling After Adding Each Virtual Machine

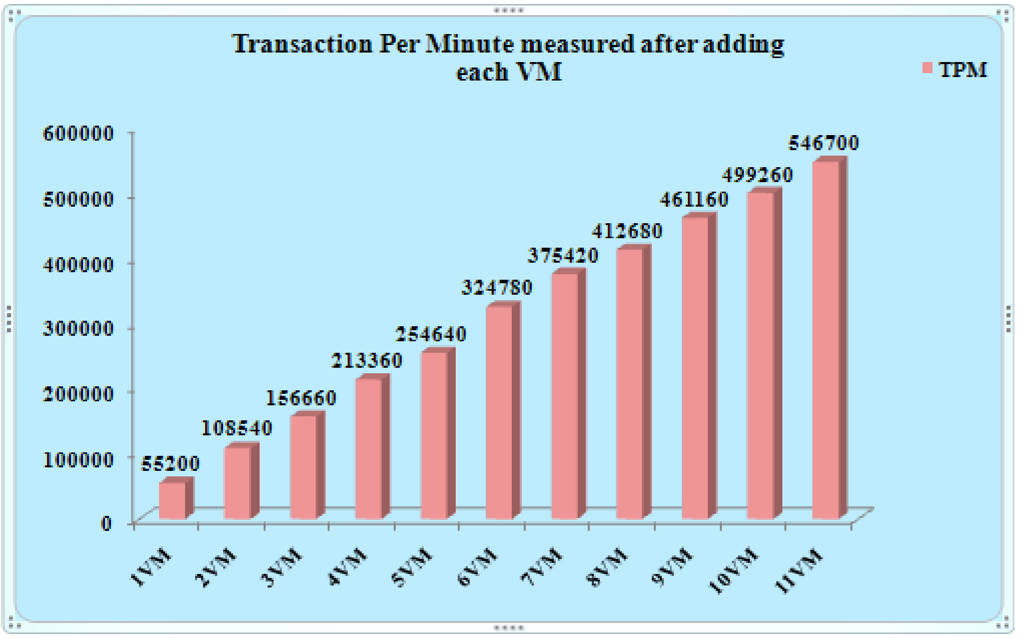


Figure 22 shows the resource Utilization of Cisco UCS B250 M2, especially CPU and Memory of Physical Machine after adding each virtual machine one after another one large database in each virtual machine and concurrently running OLTP workload one each virtual machine. Workload was capped to make sure that resources (Memory and CPU) were not saturated. In this testing process we observed 64 percent of memory utilization and 99 percent CPU Utilization.

Figure 22. CPU and Memory Utilization of Physical Machines After Adding Each Virtual Machines

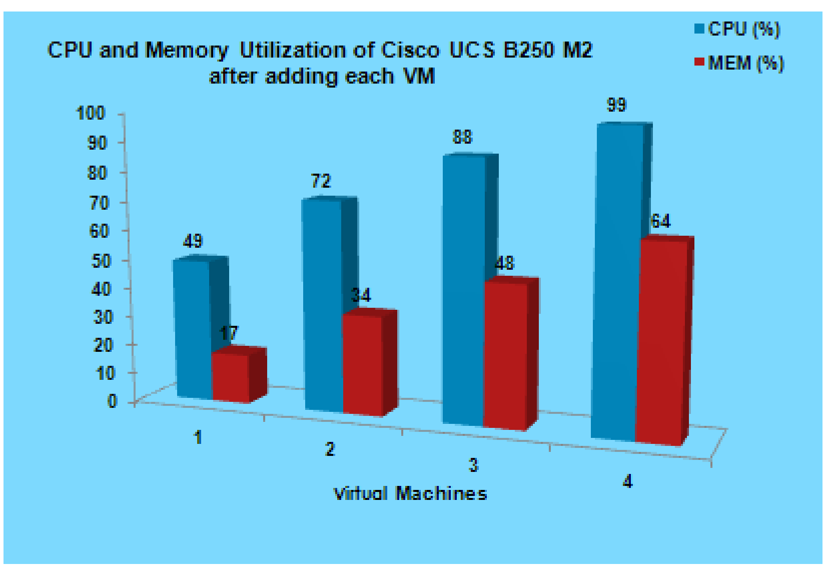


Figure 23 shows the test results of Cisco UCS B250 M2, OPM recorded after adding one large database on each virtual machine into the same physical machine. In this scenario we added one large database on each virtual machine and there were 4 virtual machines created on the same physical machine. We found linear scaling of OPM as well as the linear scaling of resource utilization.

Figure 23. OPM and Database Scaling

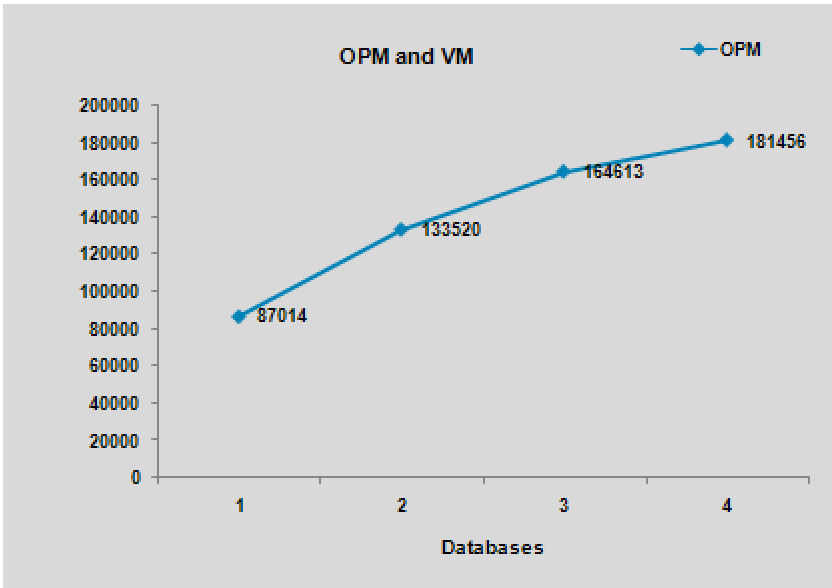


Figure 24 shows the IO characteristics of Cisco UCS B250 M2, the IOPs has been measured using EMC Unisphere monitoring GUI interface and it is recorded after adding each virtual machine into the same physical machine. In this scenario we scaled up to four virtual machines on Cisco UCS B250 M2, which typically uses 15 disks for data lun configured with RAID 5 (4+1)*3) and 8 disks for Log LUN configured with RAID 10(4+4).

Figure 24. IO Characteristics of Cisco UCS B250 M2 Measured After Adding Each Virtual Machine

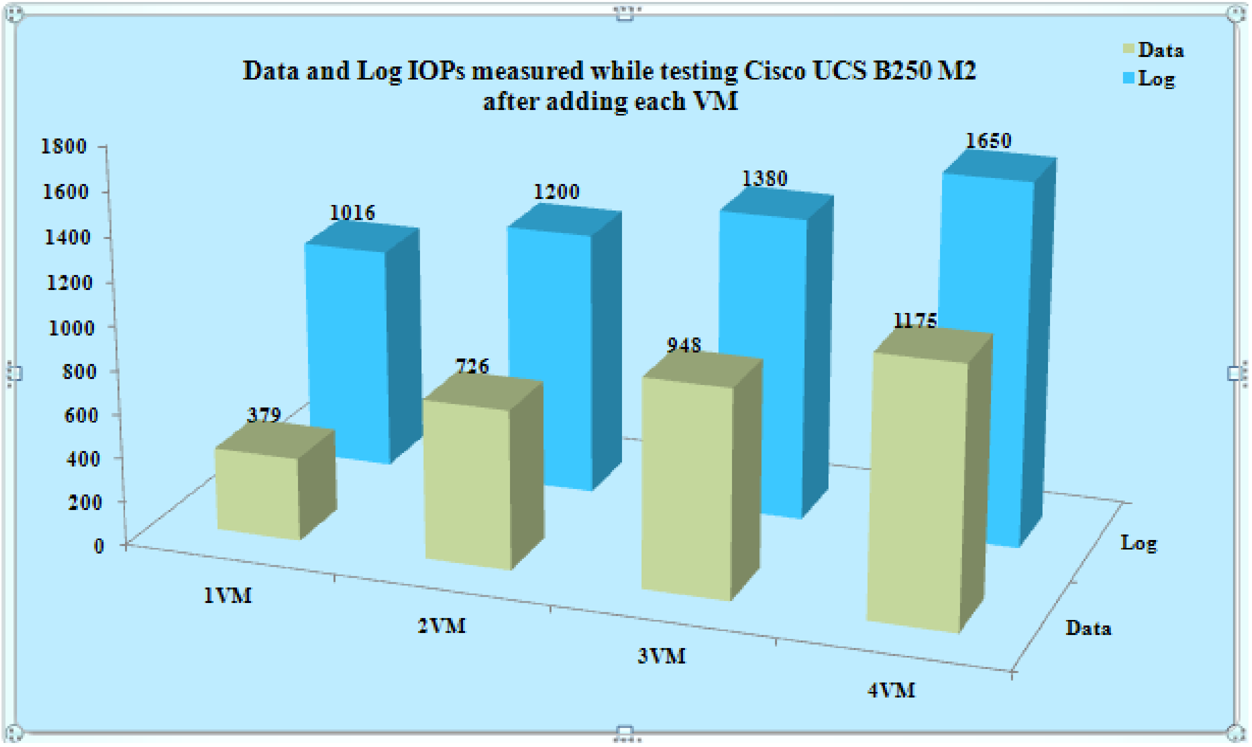
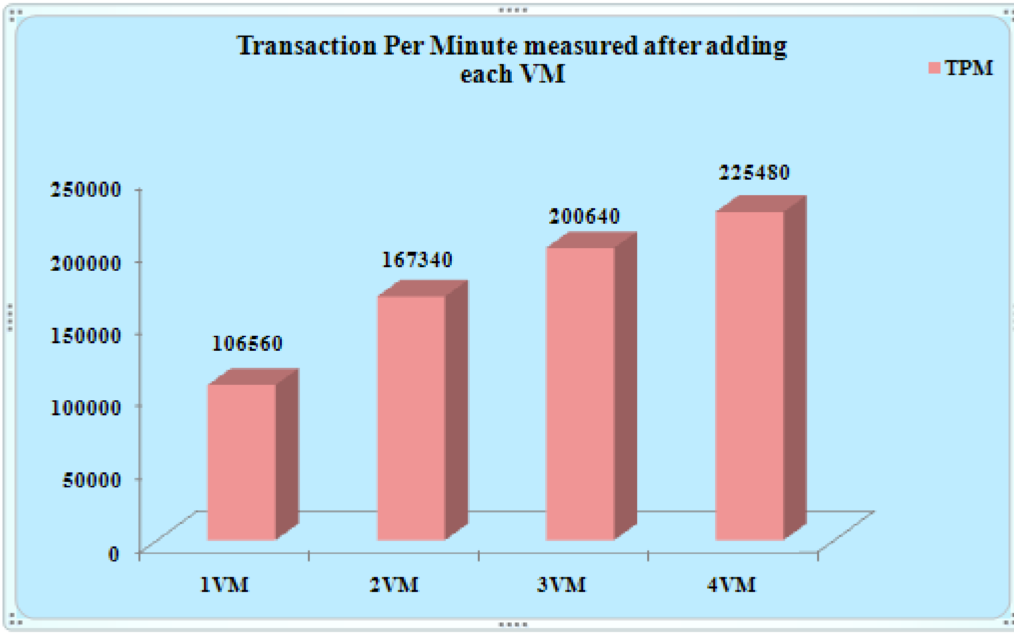


Figure 25 shows the test results of Cisco UCS B250 M2, Transaction per Minute (TPM) is measured using Performance monitor and it is recorded after each virtual machine into the same physical machine. In this scenario we scaled one large database on each virtual machine and there were four virtual machines created on the same physical machine. With the these results B250 appears to be platform of choice for entry level consolidation of large databases. It provides sufficient memory capacity as well I/O bandwidth needed for large database deployment under virtualized environment.

Figure 25. Transaction per Minute Scaling After Adding Each Virtual Machine



Scenario 3: SQL Server Consolidation Using Cisco UCS B440 M2 With 100 GB Database

As described in workload section, we started test with one large database in the SQL Server Instance running on the virtual machine. We captured the test results at the end of test. Later we keep on adding one virtual machine at a time with the same memory and database instance. At every virtual machine count we captured metrics of the workload. It was observed that the best utilization of CPU and Memory of each virtual machine as well as Physical machine was sustained until 7 virtual machines were deployed on the same physical machine.

Figure 26 shows the resource Utilization of Cisco UCS B440 M2, especially CPU and memory of physical machine after adding each virtual machine one after another with one large database in each virtual machine and concurrently running OLTP workload one each virtual machine. Workload was capped to make sure that resources (Memory and CPU) were not saturated. In this testing process we observed 86 percent of memory utilization and 74 percent CPU Utilization. As it is shown in the graph, physical machine still had enough CPU and memory to handle any unexpected spikes in the workloads.

Figure 26. CPU and Memory Utilizaiton of Physical Machines After Adding Each Virtual Machine

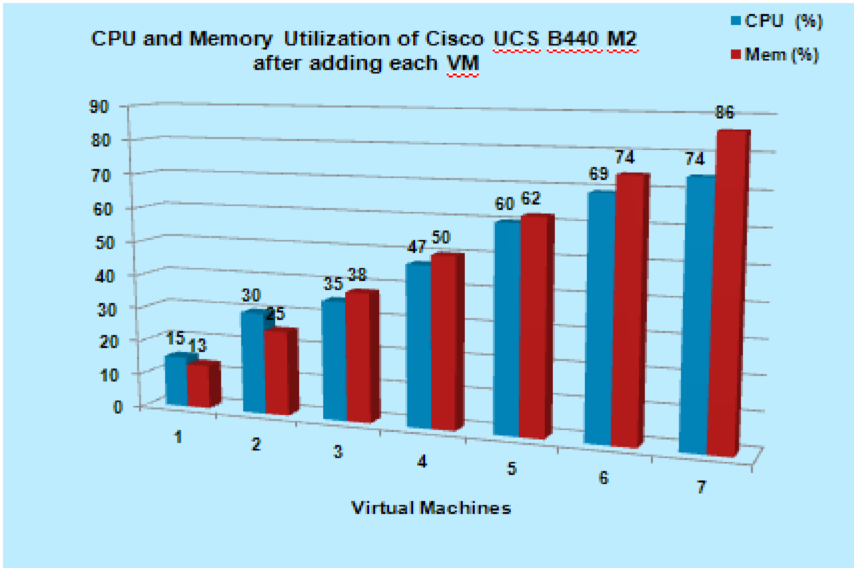


Figure 27 shows the test results of Cisco UCS B440 M2, OPM recorded after adding one large database on each virtual machine into the same physical machine. In this scenario we added one large database in each virtual machine and there were seven virtual machines created on the same physical machine. We found linear scaling of OPM as well as the linear scaling of resource utilization.

Figure 27. OPM and Database Scaling

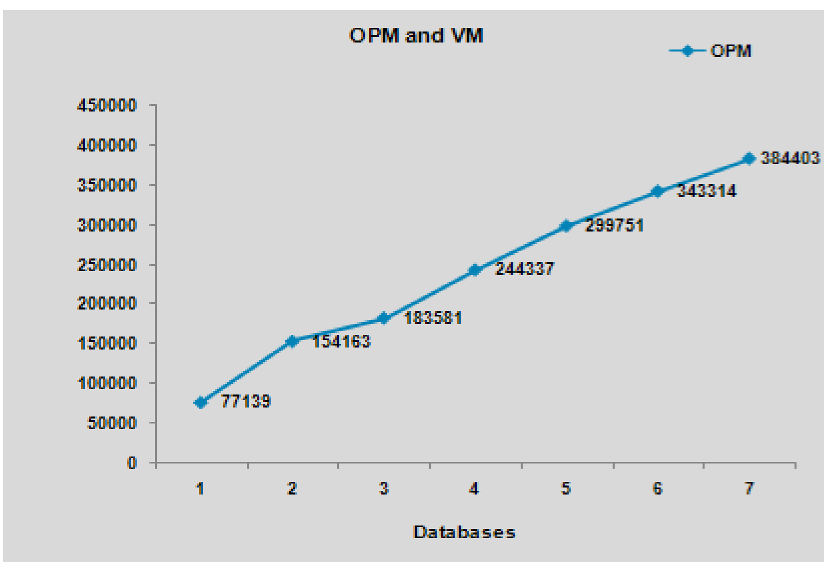


Figure 28 shows the IO characteristics of Cisco UCS B440 M2, the IOPs has been measured using EMC Unisphere monitoring GUI interface and it is recorded after adding each virtual machine into the same physical machine. In this scenario we scaled up to 7 virtual machines on Cisco UCS B440 M2, which typically uses 20 disks for data lun configured with RAID 5 (4+1)*4) and 8 disks for Log LUN configured with RAID 10(4+4).

Figure 28. IO Characteristics of Cisco UCS B440 M2 Measured After Adding Each Virtual Machine

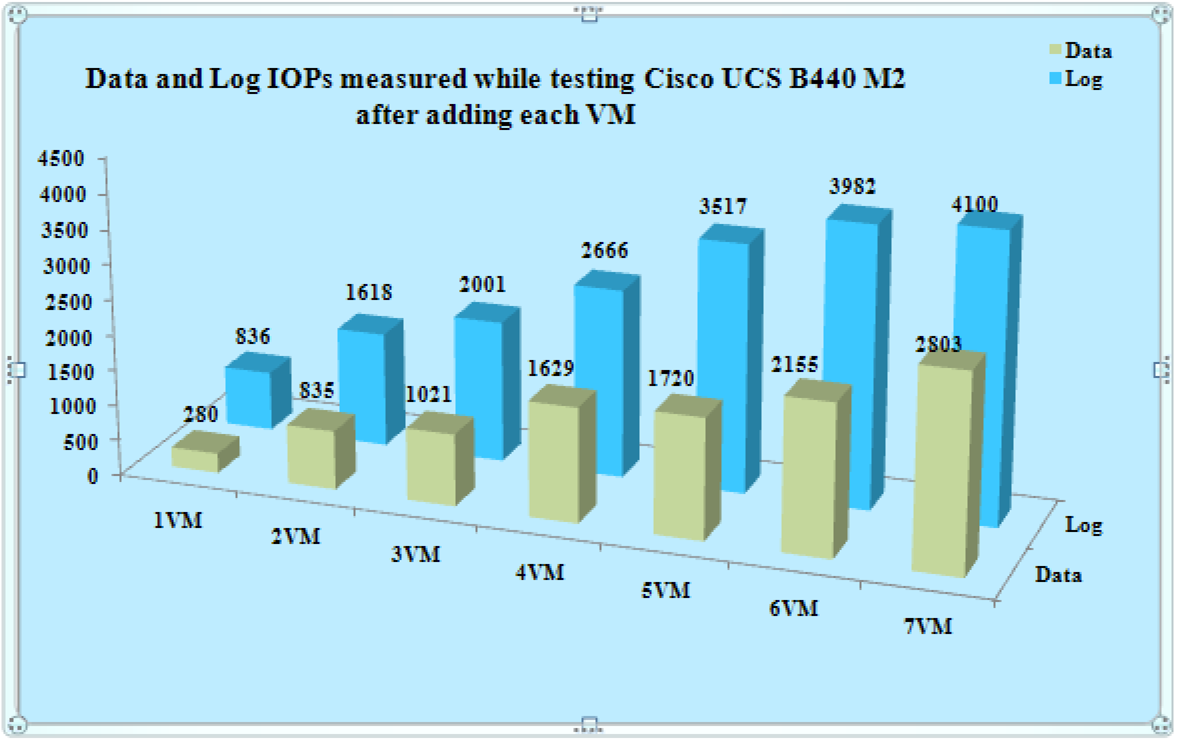
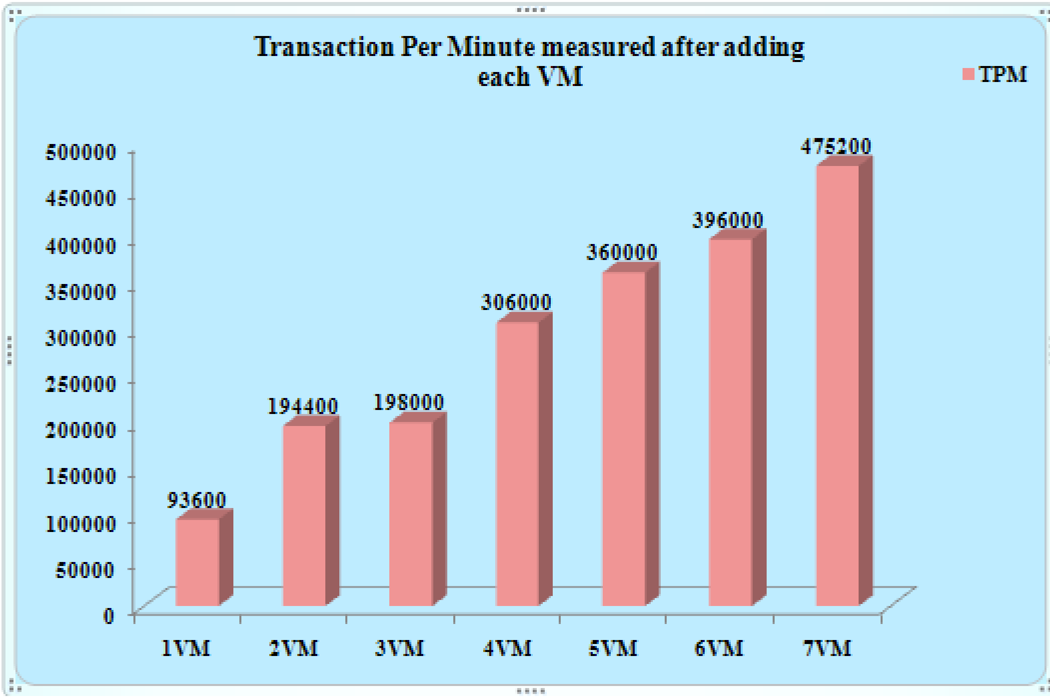


Figure 29 shows the test results of Cisco UCS B440 M2, Transaction per Minute (TPM) is measured using Performance monitor and it is recorded after each virtual machine into the same physical machine. In this scenario we scaled one large database on each virtual machine and there were seven virtual machines created on the same physical machine. With these results it is evident that B440-M2 offers the best infrastructure for Large database consolidation on virtualized environment where demand for memory, compute and I/O are much higher, typical of enterprise database backend deployments.

Figure 29. Transactions per Minute Scaling After Adding Each Virtual Machine



Consolidated OPM, CPU and Memory Usage with Small Database Consolidation

Cisco UCS B200 M2 Blade Server with Intel Xeon 5690 and 192 GB of RAM					
Number of VM's	OPM	TPM	CPU Util	Mem Util	Total IOPs
2	48408	108000	18%	33%	1677
4	97027	216000	36%	66%	3403
6	138576	295200	67%	93%	4778
Cisco UCS B230 M2 Blade Server with Intel Xeon 4870 and 256 GB of RAM					
Number of VM's	OPM	TPM	CPU Util	Mem Util	Total IOPs
2	48461	108000	14%	26%	1509
4	96903	216000	21%	38%	2986
6	143173	316800	43%	74%	4409
7	167913	374400	50%	87%	5144
Cisco UCS B250 M2 Blade Server with Intel Xeon 5690 and 384 GB of RAM					
Number of VM's	OPM	TPM	CPU Util	Mem Util	Total IOPs
2	47738	108540	19%	17%	1468
4	95054	213360	42%	33%	2806
6	141716	324780	57%	53%	4439
8	190726	412680	84%	68%	5900
10	233279	499260	94%	80%	7439
11	252718	546700	98%	88%	7775

Based on the results we can clearly identify the boundaries of various deployment scenarios that one encounters consolidating Microsoft SQL server on a virtualized environment. Following table describes generic guideline for picking suitable blade while consolidating SQL server on virtualized environment. For an entry level configuration B200-M2 provides an ideal platform for consolidating up to 24 databases. As memory requirement increases one can choose B230-M2 as it provides larger memory capacity with higher computing power. Note that i/o capacity was not saturated in any of the experiments and we still have lot of room from I/O capacity. With larger I/O requirement and better redundancy B250-M2 becomes the choice for the consolidation with its good balance of memory and I/O bandwidth needed for Microsoft SQL server. Note that in this paper testing was done with 256G on B230-M2. As seen in from the result, B230-M2 shows linear scalability in terms of number virtual machines and can potentially support close to 17 virtual machines hosting around 68 databases.

Workload Characteristics	I/O Characteristics	Suggested Blade	Max Memory
Average memory requirement	Average i/o	B200-M2	192G
Highly memory bound	Average i/o	B230-M2	512G
Moderately memory bound	High i/o	B250-M2	384G

Consolidated Results of OPM, CPU and Memory Usage with Large Database Consolidation

Cisco UCS B250 M2 Blade Server with Intel Xeon 5690 and 384 GB of RAM					
Number of VM's	OPM	TPM	CPU Util	Mem Util	Total IOPs
2	133520	167340	72%	34%	1926
3	164613	200640	88%	48%	2328
4	181456	225480	99%	64%	2825
Cisco UCS B440 M2 Blade Server with Intel Xeon 4870 and 512 GB of RAM					
Number of VM's	OPM	TPM	CPU Util	Mem Util	Total IOPs
2	154163	194400	30%	25%	2453
4	244337	306000	47%	50%	4295
6	343314	396000	69%	74%	6137
7	384403	475200	74%	86%	6903

On the large database (100Gb) consolidation scenario, B250-M2 acts as entry point with its balanced memory and i/o bandwidth. As the CPU and memory requirement increases, B440-M2 provides the ideal platform to handle large workload and IOPs. As mentioned before, IOPS shown here are not the max that either B250 or B440 can handle. I/O limits were never touched in this test and there is enough headroom. Following table summarizes the choices blades based on the workload.

Workload Characteristics	I/O Characteristics	Suggested Blade	Max Memory
Moderately memory requirement	Average i/o	B250-M2	384G
Highly memory bound	High i/o	B440-M2	512G

Conclusion

This study provides an overview of considerations to develop a strategy to maximize hardware utilization and reducing costs associated with database sprawl. It describes the strategies and reference architectures as a starting point to consolidate SQL Server on the Cisco Blade server using a building block approach to design, configure, and deploy using best practice recommendations to simplify IT.

To simplify the design and deployment of a virtualized infrastructure, Cisco offers Solution Architectures bundles for blade servers, VMware vSphere. The bundles provides configuration and best practices to achieve full redundancy—with no single point of failure, scalability, and ease of management.

The tests performed at Cisco labs showed that significant gains can be achieved by developing a strategy to maximize hardware utilization, reducing sprawl, power and cooling costs by consolidating and vitalizing SQL Server on the latest Cisco blade servers, while providing the performance to meet the most demanding customers' workloads.

For More Information

For more information about the Cisco and Microsoft alliance, visit www.cisco.com/ go/microsoftalliance or e-mail cisco_microsoft_info@cisco.com.

Acknowledgments

This document was made possible through the efforts of Vijayakumar Durairaj, with thanks to Vadiraja Bhatt.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C07-614438-01 08/11